# Patient customer feedback summarization for hospital management system

Course Name: Applied Machine Learning for Text Analysis

Team Members:

2300080176 - K Shiva

2300033358 – K Santosh

# Abstract

In the healthcare sector, patient feedback plays a vital role in assessing service quality and improving hospital operations. However, manually reviewing and analyzing a large volume of feedback from patients is time-consuming and often leads to inconsistent conclusions. This project, titled "Patient Customer Feedback Summarization for Hospital Management System," aims to automate the process of collecting, analyzing, and summarizing patient opinions using natural language processing and machine learning techniques.The system processes patient comments collected through hospital portals or surveys, performs text preprocessing such as tokenization, stopword removal, and lemmatization, and applies algorithms like TF-IDF and sentiment analysis to identify key insights. Machine learning models are then used to classify feedback as positive, negative, or neutral, and generate concise summaries that capture the main points of patient experience.This automated feedback summarization system enables hospital management to quickly understand patient satisfaction levels, detect recurring issues, and take corrective measures efficiently. By converting unstructured feedback into meaningful summaries, the system supports data-driven decision-making and helps improve the quality of healthcare services and overall patient experience.

# Introduction

In today's healthcare environment, patient satisfaction has become a key factor in evaluating the performance of hospitals and healthcare institutions. Patients often share feedback about their experiences related to medical services, staff behavior, waiting times, billing, cleanliness, and overall care quality. This feedback contains valuable information that can help hospital management understand patient expectations and identify areas for improvement.However, manually analyzing a large number of feedback entries is a challenging and time-consuming task.To overcome this issue, automated feedback summarization systems can be used to collect and analyze patient comments efficiently. By using natural language processing and machine learning techniques, patient opinions can be processed to extract key insights and summarize the overall sentiment. The system identifies common topics such as doctor consultation, nursing care, and facility management while classifying feedback as positive, negative, or neutral.The Patient Customer Feedback Summarization system for the Hospital Management System aims to simplify feedback analysis and provide meaningful summaries that support data-driven decision-making. It helps hospital administrators monitor service quality, detect recurring problems, and take necessary actions to enhance the patient experience. This approach not only saves time and effort but also ensures that every patient's voice contributes to the continuous improvement of healthcare services.

## Literature Review

Patient feedback analysis has gained increasing attention in recent years as hospitals and healthcare institutions strive to improve service quality and patient satisfaction. Several studies have explored the use of natural language processing (NLP) and machine learning (ML) techniques to automatically analyze and summarize textual feedback from patients. Early research primarily focused on manual or rule-based methods to identify keywords and sentiments in patient comments, but these approaches were limited in accuracy and scalability.Recent advancements in text mining have introduced machine learning models capable of handling large and complex datasets. Techniques such as Bag-of-Words, TF-IDF, and sentiment analysis have been widely applied to classify feedback as positive, negative, or neutral. Researchers have also used clustering algorithms to group similar complaints and highlight frequently mentioned issues such as long waiting times or staff behavior. These approaches help in converting unstructured patient opinions into actionable insights for hospital management.

## Flow Diagram

The following diagram illustrates the process flow of the Customer Feedback Summarization system
1. Data Collection → 2. Data Preprocessing → 3. Feature Extraction (TF-IDF) → 4. Model Training (Naive Bayes) → 5. Evaluation → 6. Generation of final report.

## Methodology

The methodology for Patient Customer Feedback Summarization for Hospital Management System involves several systematic steps to process and analyze patient opinions effectively. The process begins with data collection, where feedback is gathered from hospital management systems, online surveys, or feedback forms. The collected data is stored in CSV format containing textual feedback and corresponding sentiment labels such as positive, negative, or neutral.Next, data preprocessing is performed to clean and prepare the text for analysis. This step includes removing special characters, punctuation, and stopwords, converting all text to lowercase, and tokenizing the words. These preprocessing tasks help ensure that the data is consistent and suitable for feature extraction. After cleaning, the Term Frequency–Inverse Document Frequency (TF-IDF) method is applied to convert the textual data into numerical vectors. TF-IDF identifies the most important words in the feedback based on their frequency and relevance, which helps the system focus on meaningful terms.Once the data is transformed into numerical features, sentiment analysis and classification are carried out using machine learning algorithms such as Multinomial Naive Bayes or Logistic Regression. The model learns patterns from the feedback text and classifies each entry into categories such as positive, negative, or neutral. After classification, text summarization techniques are applied to generate concise summaries that highlight the most important aspects of patient opinions, such as satisfaction with doctors, service quality, waiting time, and hospital facilities.

## Code and Results

The project is implemented in Python using libraries such as pandas, scikit-learn, and NumPy. These libraries are used for handling data, preprocessing text, extracting features, and training the machine learning model

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report

#LOAD DATASET

df = pd.read_csv('patient_feedback.csv')
df['Feedback'] = df['Feedback'].astype(str)

#SPLIT DATA

X_train, X_test, y_train, y_test = train_test_split(df['Feedback'],
df['Sentiment'], test_size=0.2, random_state=42)
```

```
#TF-IDFVECTORIZATION

tfidf = TfidfVectorizer(stop_words='english')
X_train_tfidf = tfidf.fit_transform(X_train)
X_test_tfidf = tfidf.transform(X_test)

#TRAIN AND EVALUTE

model = MultinomialNB()
model.fit(X_train_tfidf, y_train)
y_pred = model.predict(X_test_tfidf)
print('Accuracy:', accuracy_score(y_test, y_pred))
print('\nClassification Report:\n', classification_report(y_test,
y_pred))
```

## Limitations

Although the Patient Customer Feedback Summarization system provides an efficient way to analyze and summarize patient feedback, it has certain limitations. The performance of the model largely depends on the quality, quantity, and diversity of the dataset used for training. If the dataset is small, unbalanced, or contains biased information, the accuracy of the sentiment classification and summarization may be affected. The system also focuses mainly on textual feedback and does not consider other forms of data such as ratings, timestamps, or demographic information that could provide deeper insights into patient satisfaction.Another limitation is that traditional models like Naive Bayes or Logistic Regression may not fully capture the complex meanings, emotions, and context of patient language, especially in cases where feedback is written informally or contains medical terminology. The model might also misclassify feedback that contains mixed sentiments, such as when a patient praises one service but criticizes another in the same comment.

## Analysis

The analysis of the Patient Customer Feedback Summarization system focuses on evaluating the model's performance and understanding how effectively it classifies and summarizes patient feedback. The Multinomial Naive Bayes classifier, trained on TF-IDF features, has shown consistent accuracy in identifying the sentiment of patient comments. It performs particularly well in detecting clear positive or negative expressions, such as satisfaction with staff behavior or dissatisfaction with waiting time. The model demonstrates good precision and recall across most sentiment categories, indicating that it can successfully capture the overall tone of patient feedback.The classification report and confusion matrix reveal that while the model performs well for feedback with strong emotional cues, it occasionally struggles with neutral or mixed comments that contain both positive and negative opinions. For instance, a patient might express appreciation for medical care but complain about billing issues, leading to potential ambiguity in classification. Despite these challenges, the system achieves a satisfactory level of accuracy and provides reliable results when tested on real-world feedback data.

# Conclusion

The Patient Customer Feedback Summarization system demonstrates the effective use of natural language processing and machine learning techniques in analyzing and summarizing patient opinions. By applying text preprocessing, TF-IDF feature extraction, and sentiment classification, the system is able to automatically process large volumes of patient feedback and generate concise summaries that highlight the main points. The use of a Multinomial Naive Bayes model provides a reliable and efficient approach for sentiment classification, allowing hospital management to quickly understand the overall patient experience and satisfaction level.This automated system helps in reducing the manual effort required to review and categorize patient feedback, making the process faster and more consistent. It assists administrators in identifying areas that require improvement, such as waiting times, staff behavior, or service quality, while also recognizing the aspects that patients appreciate. The insights generated from the system support data-driven decision-making, leading to better management strategies and improved patient care.Although the system performs well, future work can focus on incorporating more advanced deep learning models such as LSTM or BERT to improve the accuracy of sentiment detection and contextual understanding. Integrating real-time feedback analysis and visual dashboards could further enhance its usability. Overall, the project contributes to the development of intelligent hospital management solutions that aim to improve healthcare quality and strengthen the relationship between patients and medical service providers.